

Inteligencia artificial y derechos humanos: entre oportunidades, riesgos y remedios

*Artificial intelligence and human rights:
between opportunities, risks and remedies*

*Inteligência artificial e direitos humanos:
entre oportunidades, riscos e remédios*

María Fernanda Sánchez Díaz¹

Universidad Nacional Autónoma de México (México)

Orcid: <https://orcid.org/0000-0002-7519-5258>

RESUMEN: La inteligencia artificial representa uno de los avances tecnológicos más importantes para la humanidad; sin embargo, su evolución no ha estado acompañada de un marco normativo que establezca las bases para su diseño y desarrollo, lo que puede derivar en afectaciones a los derechos humanos, al incrementar los niveles de vulnerabilidad y disminuir su eficacia. La inteligencia artificial avanza a un ritmo que la normativa difícilmente podrá contener, por lo que resulta fundamental sentar las bases jurídicas desde una perspectiva de derechos humanos, a fin de prevenir sesgos algorítmicos que fomenten la discriminación, la desigualdad y la violencia. No debemos pensar en la inteligencia artificial como un enemigo a vencer, sino como una herramienta que puede convertirse en el mejor aliado de quienes sepan utilizarla y aprovecharla para potencializar sus capacidades. La elaboración del presente trabajo se desarrolló mediante un análisis documental de corte cualitativo, dirigido a examinar de manera crítica las transformaciones tecnológicas asociadas a la inteligencia artificial y su impacto en los derechos humanos. En este contexto, se revisaron instrumentos jurídicos, literatura especializada, casos recientes y regulaciones emergentes, con la finalidad de identificar patrones, tensiones y vacíos que evidencian los desafíos actuales en materia de gobernanza algorítmica. Este enfoque permitió evaluar las fuentes seleccionadas en un marco interpretativo sólido, que fuera capaz de mostrar cómo la aceleración tecnológica exige una reflexión profunda y un replanteamiento normativo que asegure la protección efectiva de la dignidad humana.

PALABRAS CLAVE: inteligencia artificial; derechos humanos; sesgos; algoritmos; potencializador.

¹ Doctora en Derecho por la Universidad Nacional Autónoma de México. Máster en Protección Internacional de los Derechos Humanos por la Universidad de Alcalá (España). Catedrática de la Universidad Nacional Autónoma de México.

ABSTRACT: Artificial intelligence represents one of the most important technological advancements for humanity; however, its development has not been accompanied by a regulatory framework that establishes the foundations for its design and evolution, which may lead to human rights violations by increasing vulnerability and reducing their effectiveness. Artificial intelligence advances at a pace that regulation can hardly keep up with, making it essential to establish legal foundations from a human rights perspective in order to prevent algorithmic biases that foster discrimination, inequality, and violence. We should not view artificial intelligence as an enemy to defeat, but rather as a tool that can become the best ally for those who know how to use and leverage it to enhance their capabilities. This article was developed through a qualitative documentary analysis grounded in a critical review of regulatory frameworks, scholarly literature, and recent cases concerning the development of artificial intelligence. This methodological approach enabled the identification of key risks, regulatory gaps, and ethical tensions affecting the protection of human rights, providing a structured interpretation of the current landscape and the challenges posed by algorithmic governance.

KEYWORDS: artificial intelligence; human rights; biases; algorithms; enhancer.

RESUMO: A inteligência artificial representa um dos avanços tecnológicos mais importantes para a humanidade. No entanto, sua evolução não foi acompanhada por um marco normativo que estabeleça as bases para sua criação e desenvolvimento, o que pode resultar em violações de direitos humanos, aumentando os níveis de vulnerabilidade e reduzindo sua efetividade. A inteligência artificial avança de forma tão acelerada que a normatização dificilmente conseguirá acompanhá-la. Por isso, é fundamental estabelecer fundamentos jurídicos com enfoque nos direitos humanos, a fim de evitar vieses algorítmicos que promovam discriminação, desigualdade e violência. A inteligência artificial não deve ser vista como um inimigo a ser combatido, mas sim como uma ferramenta que pode ser a melhor aliada daqueles que souberem utilizá-la e aproveitá-la para potencializar suas capacidades. Este artigo foi elaborado por meio de uma análise documental de caráter qualitativo, fundamentada na revisão crítica de marcos regulatórios, literatura especializada e casos recentes relacionados ao desenvolvimento da inteligência artificial. Essa abordagem metodológica permitiu identificar os principais riscos, lacunas regulatórias e tensões éticas que afetam a proteção dos direitos humanos, oferecendo uma interpretação estruturada do panorama atual e dos desafios impostos pela governança algorítmica.

PALAVRAS-CHAVE: inteligência artificial; direitos humanos; vieses; algoritmos; potencializador.

SUMARIO: 1 Inteligencia artificial y derechos humanos: el dilema del siglo XXI. 2 De la Declaración Universal al AI Act: el derecho frente a la revolución algorítmica. 3 Gobernanza algorítmica con enfoque en derechos: sesgos, regulación y participación ciudadana. 4 A modo de cierre. Referencias.

1 Inteligencia artificial y derechos humanos: el dilema del siglo XXI

La inteligencia artificial (IA) se ha consolidado como uno de los desarrollos tecnológicos más disruptivos del siglo XXI, desafiando marcos jurídicos, éticos y sociales preexistentes. En este contexto, resulta urgente reflexionar sobre su impacto en los derechos humanos y analizar si su evolución, sin una regulación eficaz, podría profundizar desigualdades estructurales o vulnerar principios fundamentales como la dignidad, la privacidad o la igualdad.

No obstante, los orígenes de esta tecnología se remontan a mediados del siglo XX, con figuras clave como Alan Turing, considerado el padre de esta disciplina. Durante la Segunda Guerra Mundial (García, 2017), Turing diseñó el sistema para descifrar el "Código Enigma" del ejército alemán, lo que contribuyó significativamente a acortar la duración del conflicto bélica.

A partir de este logro, surgió una pregunta central que aún resuena: ¿pueden las máquinas pensar? Esta inquietud motivó a Turing a proponer la famosa "prueba de Turing", destinada a determinar si una persona puede distinguir, sólo a través del lenguaje, si interactúa con un ser humano o con una máquina. En su artículo Fundamental, publicado en 1950, Turing anticipó:

[...] <<[hacia el año 2000] será posible programar computadoras [...] [que] dominen la imitación tan bien que un interrogador promedio no tendrá más del 70% de probabilidades de acertar tras cinco minutos de interrogatorio>>
[...] Con <<identificación correcta>>, Turing se refería a distinguir con precisión entre respuestas textuales generadas por humanos y generadas por computadora. (Verma, 2024)

La inteligencia artificial surgió, en ese contexto, como una solución técnica frente a un problema que cobraba millones de vidas humanas. En sus orígenes, representó una promesa de innovación científica y tecnológica. No obstante, su desarrollo acelerado exige hoy un análisis riguroso desde la óptica de los derechos humanos, a fin de valorar si esta tecnología contribuye efectivamente al bien común o, por el contrario, genera nuevas formas de vulnerabilidad y exclusión.

Así como en su momento la IA sirvió para salvar vidas, hoy debemos interrogarnos si su diseño, entrenamiento y aplicación pueden causar más daño que los beneficios que inicialmente generó. Esta preocupación cobra particular relevancia cuando los sistemas algorítmicos son entrenados con datos sesgados o estructurados sin considerar su impacto diferencial en poblaciones vulnerables.

Actualmente, en pleno siglo XXI, y con el acceso casi universal a tecnologías basadas en inteligencia artificial, estas comienzan a simular de forma cada vez más precisa el pensamiento y la comunicación humana. Con instrucciones adecuadas, contexto suficiente y arquitecturas diseñadas ad hoc, los modelos de lenguaje y los sistemas generativos son capaces de imitar la forma de expresión de una persona. Esta capacidad, conocida como atribución de tono de voz en la generación textual, también se extiende a la clonación de

imagen y voz. Si bien estos resultados aún presentan imperfecciones, la realidad es que nos acercamos a un nivel de sofisticación tal que dificulta la distinción entre lo real y lo artificial.

Esta confusión plantea riesgos significativos desde la perspectiva de los derechos humanos. La manipulación de contenido, las suplantaciones de identidad y la generación de información falsa son solo algunos ejemplos del uso potencialmente abusivo de estas tecnologías en contextos sensibles.

Este panorama ha puesto sobre la mesa el riesgo que enfrentan los derechos fundamentales ante la falta, o insuficiencia, de una regulación eficaz. Si bien se han producido avances normativos en algunas regiones del mundo, como en la Unión Europea con el AI Act, o más recientemente en el estado de California, Estados Unidos, el escenario global aún es incipiente y fragmentado.

En este contexto, resulta fundamental analizar si debe priorizarse el desarrollo tecnológico sobre los derechos fundamentales, o bien, si es legítimo el uso de la inteligencia artificial en áreas como la seguridad pública, la administración de justicia, la educación o la salud sin una regulación robusta. Del mismo modo, es urgente considerar cómo equilibrar la innovación tecnológica con principios esenciales como la igualdad, la privacidad y la dignidad humana, siendo estos últimos los que enfrentan mayores riesgos de afectación en el corto plazo.

Como se abordará más adelante en esta investigación, el desarrollo de sistemas de inteligencia artificial requiere grandes volúmenes de datos que, en muchos casos, no están exentos de reproducir sesgos estructurales. Esta problemática se agrava cuando los algoritmos operan bajo criterios opacos, o cuando los conjuntos de datos no han sido diseñados con una perspectiva de inclusión, equidad o representatividad. La discriminación algorítmica, por tanto, exige un análisis profundo desde el enfoque de derechos humanos, especialmente en lo relativo al principio de igualdad ante la ley, el debido proceso y la protección de datos personales, particularmente datos sensibles.

La inteligencia artificial debe concebirse como una herramienta para potenciar el desarrollo humano, no como un mecanismo de reproducción de desigualdades ni como instrumento de control social. No es la tecnología la que vulnera derechos, sino su diseño, implementación y uso sin perspectiva jurídica, ética ni social.

Además, la inteligencia artificial representa uno de los desafíos normativos más relevantes de nuestra era. Se trata de una tecnología que avanza más rápido que los marcos regulatorios que pretenden disciplinarla, lo cual nos obliga a replantear una de las preguntas centrales de nuestro tiempo: ¿cómo garantizar la eficacia de los derechos humanos en un contexto donde los algoritmos toman decisiones que afectan nuestra privacidad, dignidad y acceso a oportunidades.

En 2023, tras la aparición de ChatGPT como una herramienta accesible a nivel global, diversos expertos firmaron una carta abierta advirtiendo sobre los riesgos del desarrollo acelerado de la inteligencia artificial. En ella, señalaron que los laboratorios dedicados a esta tecnología estaban inmersos en “una carrera fuera de control para desarrollar e implementar mentes digitales cada

vez más poderosas, que nadie, ni siquiera sus creadores, puede comprender, predecir o controlar de forma fiable”, La declaración fue firmada por más de mil personas, entre ellas Elon Musk, el cofundador de Apple, Steve Wozniak, y representantes de la firma DeepMind, entre otros (BBC, 2023).

No obstante, resulta relevante señalar que, pese a firmar dicha advertencia, Elon Musk lanzó en noviembre de 2023 su propia inteligencia artificial denominada “Grok” (UNADE, 2025), integrada en su red social “X”. Esta herramienta, desarrollada por xAI, tiene la capacidad de acceder en tiempo real a publicaciones, tendencias y conversaciones dentro de la plataforma. La contradicción entre el llamado a frenar el avance tecnológico y el lanzamiento de una IA con semejantes capacidades plantea interrogantes legítimos sobre la verdadera motivación de algunos actores en esta discusión: ¿se trata de una preocupación genuina por el bienestar colectivo o de una estrategia al control de mercado, poder e influencia?

Junto con ello, cabe preguntarse cómo prevenir el uso indebido de estas herramientas cuando aún no contamos con una cultura ética consolidada para su manejo. Lejos de una utilización responsable, se han documentado casos donde estas tecnologías han sido empleadas para vulnerar el honor, la imagen y la intimidad de las personas. Un ejemplo claro son las llamadas falsificaciones profundas²: videos que muestran imágenes falsas, habitualmente del rostro de una persona, que parecen reales y que han sido generadas mediante IA, muchas veces con fines pornográficos (Visus, 2021) o de extorsión. Estas prácticas generan un daño moral y psicosocial profundo, cuyas consecuencias a corto, mediano y largo plazo aún no alcanzamos a dimensionar en toda su magnitud.

Frente a este escenario, resulta indispensable avanzar hacia una regulación que no sea meramente reactiva, sino preventiva y centrada en la protección de los derechos fundamentales. La inteligencia artificial no puede quedar al margen del derecho. Por el contrario, su desarrollo debe enmarcarse en los principios de justicia social, transparencia y dignidad humana.

En suma, la inteligencia artificial representa un panorama sumamente complejo para los derechos humanos en el siglo XXI. Si bien sus orígenes están ligado al ingenio humano, también es cierto que no siempre ha sido concebida con fines humanitarios, sino en muchos casos con propósitos económicos y comerciales. Su desarrollo actual plantea dilemas éticos, técnicos y, por supuesto, jurídicos de alta complejidad, que constituyen un desafío para todos los países en materia de regulación y eficacia de los derechos fundamentales.

Es menester comprender y asimilar, tanto a nivel social como individual, que la innovación tecnológica no se detendrá. En consecuencia, resulta imperativo construir un marco normativo sólido, transversal y garantista, que no solo maximice la protección de los derechos humanos, sino que también sirva de contención frente al avance acelerado de tecnologías como la inteligencia artificial. Lo que está en juego es evitar que el desarrollo tecnológico se traduzca en nuevas formas de exclusión, violencia o discriminación. Se

² Traducción propia del término “Deep Fake”.

requiere, en este sentido, alcanzar un equilibrio efectivo entre progreso tecnológico y respeto a la dignidad humana.

Bajo esta premisa, resulta necesario analizar cómo el derecho ha intentado responder a esta disrupción tecnológica. Para ello, es imprescindible revisar la evolución normativa, desde los instrumentos internacionales de protección de derechos humanos hasta las propuestas más recientes en el ámbito europeo, con particular énfasis en el denominado AI Act.

2 De la Declaración Universal al AI Act: el derecho frente a la revolución algorítmica

La irrupción de la inteligencia artificial (IA) tomó por sorpresa a las sociedades contemporáneas, y de manera aún más significativa, evidenció la falta de preparación normativa por parte de los Estados para garantizar la protección efectiva de los derechos humanos frente a los posibles usos indebidos de esta tecnología. La rápida democratización de su acceso, impulsada por el lanzamiento de plataformas como ChatGPT en 2022, marcó un punto de inflexión. En pocos meses, esta herramienta superó ampliamente la cifra de usuarios que lograron redes sociales como Facebook en sus primeras etapas, lo que puso de manifiesto no solo su potencial, sino también los retos que conlleva.

ChatGPT es un generador de lenguaje creado por la empresa estadounidense OpenAI y financiado por Microsoft. Se trata de una interfaz que permite la interacción humana con el modelo de lenguaje autorregresivo GPT (Generativa Pre-Trained Transformer), una herramienta que utiliza técnicas de inteligencia artificial, en particular el aprendizaje automatizado (machine learning), para generar textos nuevos a partir de entradas previas (Riquelme, 2023).

El impacto de la inteligencia artificial ha sido, en muchos sentidos, tan disruptivo como lo fue el acceso generalizado a Internet. Cuando ChatGPT se popularizó, existía un alto grado de desinformación, acompañado de temor en ciertos sectores ante la incertidumbre sobre el destino de los datos personales introducidos en el sistema. Aunque estos temores iniciales han disminuido con el tiempo, no deben descartarse: la confianza ciega en una tecnología sin regulación adecuada puede generar riesgos reales para la privacidad y la integridad de las personas usuarias.

Actualmente, observamos un uso creciente de herramientas de inteligencia artificial en ámbitos altamente sensibles como el derecho, la medicina o la salud mental. La preocupación no radica en su uso como apoyo, sino en su empleo sustitutivo, sin conocimiento técnico suficiente ni supervisión profesional. Se han documentado casos de despachos jurídicos que han promovido demandas con precedentes inexistentes generados por IA, así como situaciones en las que personas usuarias buscan orientación emocional mediante estas plataformas como si se tratara de terapeutas profesionales. También se han reportado casos de autodiagnósticos y automedicación basados en respuestas generadas por IA, sin considerar que estas herramientas carecen de acceso al historial clínico

completo o a la evaluación especializada que solo un profesional de la salud puede ofrecer.

La Declaración Universal de los Derechos Humanos marcó un antes y un después en el reconocimiento internacional de los derechos humanos. Surge como respuesta a los horrores vividos durante la Segunda Guerra Mundial y establece derechos fundamentales como el derecho a la vida, la igualdad, la dignidad humana, la seguridad, la libertad, el reconocimiento de la personalidad jurídica, la prohibición de tratos crueles, inhumanos o degradantes, el derecho a un recurso efectivo ante tribunales nacionales, y el derecho a la privacidad, entre otros. Esta Declaración se rige bajo los principios de universalidad, interdependencia, indivisibilidad y progresividad.

Si bien a nivel nacional los Estados ya contaban con Constituciones que reconocían derechos fundamentales, no existía un instrumento internacional que los garantizara de forma global. La Declaración Universal adquiere, por tanto, una relevancia crucial, especialmente hoy, en un mundo cada vez más interconectado donde lo que ocurre en un país puede tener repercusiones inmediatas en otros.

Setenta y siete años después de su adopción, los avances tecnológicos nos obligan a repensar la eficacia de los derechos humanos en un entorno donde la tecnología puede ser tanto una herramienta de potenciación como de vulneración masiva. La inteligencia artificial plantea un desafío reciente, pero urgente, para la comunidad internacional y los Estados en particular. Aunque puede generar beneficios en salud, educación, políticas públicas y emprendimiento, también puede ser una tecnología profundamente riesgosa si no se regula de manera adecuada.

La regulación representa un reto, precisamente por el ritmo acelerado del desarrollo tecnológico. Sin embargo, resulta indispensable establecer marcos normativos que sienten las bases para una inteligencia artificial que opere desde una perspectiva de derechos humanos.

La reciente aprobación del Reglamento de Inteligencia Artificial en la Unión Europea (AI Act) es un paso importante en este sentido. Diversas organizaciones celebraron esta iniciativa al considerar que garantiza el desarrollo de la IA respetando principios fundamentales, posicionándola como una herramienta para complementar, no reemplazar, la creatividad humana. Los firmantes de dicha declaración apoyan que el Reglamento contemple obligaciones de transparencia que permitan a los titulares de derechos conocer qué datos se están utilizando y solicitar su retiro, si así lo desean. No obstante, también se expresó preocupación por la ampliación de la excepción relativa a la minería de datos con fines comerciales, regulada en el artículo 4 de la Directiva 2019/790 (Seda, 2024).

Un aspecto clave del AI Act es la incorporación de medidas de evaluación y mitigación de riesgos, así como la exigencia de transparencia, precisión y supervisión humana constante (Seda, 2024). El papel del usuario es también central: quien utiliza una herramienta de inteligencia artificial debe estar plenamente consciente de su funcionamiento, del tipo de instrucciones que proporciona y de la necesidad de verificar los resultados obtenidos.

Por su parte, China también ha avanzado en la regulación de la inteligencia artificial. No obstante, su enfoque ha sido cuestionado, ya que no está diseñado para proteger a los usuarios, sino para incentivar el crecimiento de la industria tecnológica local:

[...] Las medidas legislativas recientes, como las medidas provisionales para regular la IA generativa y diversas leyes locales sobre IA, ofrecen poco valor protector al público chino. En cambio, estas leyes han enviado una fuerte señal de crecimiento a la industria, al tiempo que intentan coordinar a las diversas partes interesadas para acceder al progreso tecnológico. Por lo tanto, el enfoque estratégicamente laxo de China en materia de regulación puede ofrecer a sus empresas de IA una ventaja competitiva a corto plazo sobre sus homólogos europeos y estadounidenses. (Huyue, 2025)

Aunque una regulación excesiva puede obstaculizar el crecimiento económico y profesional, una regulación laxa, o peor aún, su ausencia, puede generar daños mucho más severos. El gobierno chino ha invertido enormes sumas en el desarrollo de inteligencia artificial, aunque actualmente esos presupuestos están siendo reducidos:

[...] La inversión de Alibaba promediará aproximadamente el 15% de sus ingresos durante los próximos tres años, lo que implica un total de alrededor de 71,000 millones de dólares. Esta cifra es inferior a las cantidades absolutas y relativas que Alphabet, Amazon y Microsoft invertirán cada una este año, de acuerdo con Visible Alpha. (Forbes, 2025)

En el caso de Estados Unidos, la legislación más reciente ha sido aprobada en el estado de California, Estados Unidos: la Ley de Transparencia en la Inteligencia Artificial Fronteriza (S.B. 53), firmada por el gobernador el 29 de septiembre de 2025. Para su aprobación, fue necesario llegar a un acuerdo con las empresas tecnológicas de Silicon Valley, a fin de garantizar transparencia, rendición de cuentas y seguridad en el desarrollo de sistemas de IA. La ley obliga a las empresas con ingresos superiores a los 500 millones de dólares a informar sobre el cumplimiento de normas nacionales e internacionales en materia de seguridad (Guerrero, 2025).

Las empresas tecnológicas no han estado del todo de acuerdo con este tipo de regulación, argumentando que puede frenar el desarrollo tecnológico. Por ello, algunas abogan por que la regulación de la IA sea competencia exclusiva del Congreso federal y no de los estados. La propuesta legislativa californiana promueve una inteligencia artificial "segura, ética, equitativa y sostenible".

Lo anterior evidencia que los gobiernos deberán negociar con las grandes empresas para definir el grado de control que impondrán las normas nacionales. En consecuencia, existe un riesgo latente de que la población común quede en una situación de vulnerabilidad, ya que son las personas usuarias quienes, al interactuar con estas herramientas, alimentan los sistemas de IA con sus datos personales.

Ante este panorama de transformaciones vertiginosas, el derecho enfrenta el desafío de mantenerse vigente frente al avance tecnológico que no se detiene ni se detendrá. La inteligencia artificial exige una actualización constante de los

marcos jurídicos, razón por la cual la legislación debe sentar las bases sobre las cuales debe y puede operar y evolucionar la inteligencia artificial, para efecto de que los derechos humanos no se vean en mayor grado de vulnerabilidad.

En este sentido, el AI Acta representa, el referente regulatorio para alcanzar un equilibrio entre la innovación tecnológica y la protección de la dignidad humana. La eficacia de esta disposición normativa demandará la voluntad política para hacerla exigible, y sobre todo, de la capacidad de los Estados para resistir a las presiones económicas y políticas que exijan menor control bajo el argumento del desarrollo. Como acertadamente lo señaló Victoria Martínez, gerente de Desarrollo de Negocios de Inteligencia Artificial de Red Hat en Latinoamérica:

[...] Entre 2023 y 2024 hubo muchas pruebas de concepto, pero ahora se está buscando ir a producción [...] como hoy es tan fácil construir a partir de las diversas herramientas que circulan empezaron a aflorar casos de uso sin conocimiento de las áreas de informática. Por eso es que ahora se quiere decir ordenemos esto. Si tenemos políticas de seguridad, que la IA respete las mismas políticas [...] Darle una gobernanza a todo ese cos que se ha empezado a generar [...] por el fácil uso de la tecnología. (Martínez, 2025)

Es fundamental consolidar una cultura jurídica y social en que las personas estén en el centro del diseño, implementación y evaluación de los sistemas de inteligencia artificial. Esta tecnología debe concebirse como una herramienta para potencializar las capacidades humanas, no como un mecanismo que vulnere sus derechos. Hoy en día, prácticamente todas las personas interactuamos con inteligencia artificial, de forma consciente o inconsciente, informada o desinformada, lo que impone un deber ineludible: garantizar que su desarrollo esté al servicio de la dignidad humana, y no en su contra.

3 Gobernanza algorítmica con enfoque en derechos: sesgos, regulación y participación ciudadana

Es importante comprender que la evolución de la inteligencia artificial depende de los datos que procesa, de la información que los usuarios comparten con ella y que es utilizada para su entrenamiento. Sin embargo, los algoritmos pueden estar mal diseñados desde su origen, lo que genera sesgos que necesariamente afectarán la eficiencia y la equidad de la inteligencia artificial.

Precisamente es en el ámbito de la inteligencia artificial donde se han ido detectando ciertos sesgos, como los de índole racial o de género, particularmente porque los algoritmos han sido diseñados, en su gran mayoría, por hombres de raza blanca.

Uno de los casos que se dieron a conocer en donde los algoritmos demostraron estar sesgados fue la detención de un hombre afroamericano en el estado de Louisiana, Estados Unidos: "el error de un sistema de reconocimiento facial en Estados Unidos provocó el encarcelamiento de Randall Reid, un hombre afroamericano de 28 años, al ser confundido con el autor de

dos robos de bolsos de lujo en el estado de Louisiana”, informó la agencia Associated Press (R3D, 2023).

Por su parte, ONU Mujeres ha documentado que los algoritmos de inteligencia artificial también presentan sesgos en temas de género, reforzando con ello actos de discriminación contra grupos históricamente vulnerables:

[...] En el fondo, la inteligencia artificial está compuesta de datos. Es un conjunto de tecnologías que le permite a las computadoras hacer tareas complejas más rápido que los seres humanos. Los sistemas de IA, como los modelos de aprendizaje automático, aprenden a hacer tareas a partir de los datos con que se alimentan. Cuando estos modelos se nutren de algoritmos sesgados, refuerzan las desigualdades imperantes y fomentan la discriminación de género en la IA. (Del Villar, 2025)

Los ámbitos donde se han detectado las consecuencias del sesgo en los algoritmos incluyen, por ejemplo, los procesos de contratación de personal, en los que se favorece principalmente a hombres, ya que actualmente se utilizan sistemas basados en inteligencia artificial.

Los algoritmos sesgados no solo han impactado en la reproducción de la discriminación, sino también en el fomento de actitudes violentas:

[...] Muchas empresas de redes sociales han sido criticadas por usar algoritmos, impulsados por inteligencia artificial, para determinados usuarios enviándoles contenido personalizado que reforzara sus prejuicios. Cuanto más provocador sea el contenido, más posibilidades hay de que se consuma y se comparta. (ONU, 2021)

De la misma forma en que la inteligencia artificial sesgada puede promover la violencia, también puede utilizarse con fines publicitarios, discriminando a quienes no son considerados clientes potenciales, atendiendo a sus particularidades físicas, económicas, de estatus migratorio, género o edad. Pero incluso más allá, puede convertirse en una herramienta de control social mediante la vigilancia masiva, la manipulación de la información en plataformas digitales, y la toma de decisiones automatizadas, incluyendo aquellas que impactan en procesos judiciales.

Ante este escenario, ¿qué puede hacer la población civil para prevenir ser víctima de los sesgos algorítmicos de la inteligencia artificial? Es una realidad que la IA se está introduciendo en todos los ámbitos: salud, educación, decisiones gubernamentales, economía, y también en la justicia. Una de las medidas necesarias para evitar la propagación y normalización de estos sesgos es precisamente el desarrollo de una regulación que garantice la transparencia de los sistemas. Cada vez se vuelve más urgente que el desarrollo tecnológico tenga como base transversal los derechos humanos. Es decir, que antes de aprobar cualquier avance tecnológico, se realice una evaluación de riesgo sobre su eficacia en relación con los derechos humanos.

Otro punto que puede contribuir, y que debe trabajarse de manera inmediata, es el fortalecimiento de la cultura sobre el uso de la inteligencia artificial. Se trata de una tecnología que llegó para quedarse y seguir

evolucionando. Por ello, una sociedad informada sobre los beneficios y riesgos que implica su uso, así como sobre la importancia de involucrarse activamente en la interacción con estas herramientas, podrá tener una participación más consciente. No se trata solo de emitir una instrucción a una máquina, sino de co-crear con la inteligencia artificial, de manera que esta se convierta en un potencializador de capacidades humanas, y no en una amenaza para sus usuarios, ni mucho menos para terceros.

4 A modo de cierre

El desarrollo y la disrupción de la inteligencia artificial en los distintos ámbitos ha transformado de manera profunda y vertiginosa las dinámicas sociales, económicas y, por supuesto, jurídicas del siglo XXI. Si bien esta tecnología es concebida como una herramienta poderosa, también es cierto que no ha sido bien aceptada por algunos sectores, particularmente por el desconocimiento de lo que implica su uso. En muchos casos, los temores ante lo desconocido superan la voluntad de aprender a utilizar una herramienta que podría transformar de manera positiva su entorno laboral, económico o incluso personal.

Más que tener miedo, debemos tener respeto por la tecnología. Esto implica reconocer que no podemos confiar ciegamente en ella ni entregarle toda nuestra información. Existen evidencias de sesgos en los algoritmos, así como de usos indebidos de datos personales proporcionados por los propios usuarios y utilizados para el entrenamiento continuo de los sistemas de inteligencia artificial. Sin embargo, también debemos de ser conscientes de la necesidad de involucrarnos activamente en el conocimiento de esta tecnología disruptiva y emplearla en nuestro beneficio, como una herramienta que potencializa nuestras capacidades, y no como una amenaza que inevitablemente sustituirá al ser humano.

Es cierto que algunas profesiones experimentarán transformaciones, y que ciertas actividades podrán ser reemplazadas por la inteligencia artificial. Pero esto también abre la posibilidad de que nos concentremos en aquellas tareas que realmente requieren del conocimiento y experiencia humanos. Lo que la inteligencia artificial sí sustituirá será a las personas que no aprendan a utilizarla, mientras que quienes se formen y se eduquen en su uso lograrán posicionarse con mayor ventaja.

Otro de los aspectos abordados a lo largo de este trabajo ha sido la necesidad urgente de contar con una regulación adecuada, con un enfoque transversal de derechos humanos, que funcione como base para orientar la evolución de la inteligencia artificial sin poner en riesgo la dignidad ni los derechos fundamentales. La inteligencia artificial debe consolidarse como una herramienta al servicio del ser humano, y no al revés.

Para ello, es indispensable avanzar hacia una gobernanza algorítmica con enfoque de derechos humanos, que contemple mecanismos efectivos de control, auditoría, rendición de cuentas y participación democrática. Si bien ya existen iniciativas normativas como el AI Act en la Unión Europea, las medidas

provisionales adoptadas en China o la reciente legislación en el estado de California, Estados Unidos, aún es urgente que más países se sumen a la construcción de un marco normativo que no deje esta responsabilidad en manos exclusivas de las grandes empresas tecnológicas.

La autorregulación empresarial no ofrece garantías de que se prioricen los derechos y el bienestar de las personas por encima de los intereses económicos. En este sentido, la omisión por parte de las autoridades podría tener consecuencias graves e irreversibles en materia de derechos humanos.

Referencias

BBC. La carta en la que más de 1.000 expertos piden frenar la inteligencia artificial por ser una “amenaza para la humanidad”. *BBC News Mundo*, 29 mar. 2023. Disponible en: <https://www.bbc.com/mundo/noticias-65117146>. Consultado el: 06 oct. 2025.

DEL VILLAR, Zinnya. Cómo la inteligencia artificial refuerza los sesgos de género y qué podemos hacer al respecto. *ONU Mujeres*, 05 feb. 2025. Disponible en: <https://www.unwomen.org/es/noticias/entrevista/2025/02/como-la-inteligencia-artificial-refuerza-los-sesgos-de-genero-y-que-podemos-hacer-al-respecto>. Consultado el: 06 oct. 2025.

FORBES. Los ajustados presupuestos de China para la IA se refuerzan a sí mismos; te decimos cómo. *Forbes*, 9 out. 2025. Disponible en: <https://forbes.com.mx/los-ajustados-presupuestos-de-china-para-la-ia-se-refuerzan-a-si-mismos-te-decimos-como/>. Consultado el: 06 oct. 2025.

GARCÍA, Laura. Inteligencia artificial, la herencia de Alan Turing. *Ciencia UNAM*, 03 jul. 2017. Disponible en: <https://ciencia.unam.mx/leer/631/inteligencia-artificial-la-herencia-de-alan-turing->. Consultado el: 06 oct. 2025.

GUERRERO, Manuel Alejandro. Atención con la Ley de Inteligencia Artificial de California. *El Sol de México*, 11 out. 2025. Disponible en: <https://oem.com.mx/elsoldemexico/analisis/atencion-con-la-ley-de-inteligencia-artificial-de-california-26209866>. Consultado el: 16 oct. 2025.

HUYUE ZHANG, Ángela. Las promesas y los peligros de la regulación de la inteligencia artificial en China. *Columbia Journal of Transnational Law*, n. 63, 2025. Disponible en: [https://static1.squarespace.com/static/5daf8b1ab45413657badbc03/t/67902c54fa8400aea9dc26e/1737501893309/01_CTL_63_1_Zhang+\(1\).pdf](https://static1.squarespace.com/static/5daf8b1ab45413657badbc03/t/67902c54fa8400aea9dc26e/1737501893309/01_CTL_63_1_Zhang+(1).pdf). Consultado el: 06 oct. 2025.

MARTÍNEZ, Victoria. Pasar del “caos” a la “gobernanza”: el nuevo reto en IA para Latinoamérica, según expertos. *Lópes-Dóriga Digital*, 14 out. 2025. Disponible en: <https://lopezdoriga.com/ciencia-tecnologia/pasar-del-caos-a-la-gobernanza-el-nuevo-reto-en-ia-para-latinoamerica-segun-expertos/>. Consultado el: 16 oct. 2025.

ONU. Sesgos, racismo y mentiras: afrontar las consecuencias no deseadas de la inteligencia artificial. *Naciones Unidas*, 04 ene. 2021. Disponible en: <https://news.un.org/es/story/2021/01/1485802>. Consultado el: 06 oct. 2025.

R3D. Hombre afroamericano es encarcelado en Louisiana por sesgo racial de sistema de reconocimiento facial. *Red de Defensa de los Derechos Digitales*, 05 ene. 2023. Disponible en: <https://r3d.mx/2023/01/05/hombre-afroamericano-es-encarcelado-en-louisiana-por-sesgo-racial-de-sistema-de-reconocimiento-facial/>. Consultado el: 06 oct. 2025.

RIQUELME, Rodrigo. ChatGPT hipnotiza y consterna a 6 meses de su lanzamiento. *El Economista*, 14 jun. 2023. Disponible en: <https://www.eleconomista.com.mx/tecnologia/ChatGPT-hipnotiza-y-consterna-a-6-meses-de-su-lanzamiento-20230614-0045.html>. Consultado el: 06 oct. 2025.

SEDA. La Unión Europea aprueba la Ley de IA, primer marco legal del mundo sobre Inteligencia Artificial. *Sociedad Española de Derechos de Autor*, 07 jun. 2024. Disponible en: https://sedamusica.es/noticias/noticia_ue_aprueba_ley_ia.html. Consultado el: 06 oct. 2025.

UNADE. Grok: el nuevo asistente de IA de X (antes Twitter). *Universidad Americana de Europa*, 14 abr. 2025. Disponible en: <https://unade.edu.mx/grok-la-inteligencia-artificial-de-x/>. Consultado el: 06 oct. 2025.

VERMA, Nitin. Have we passed the Turing Test, and should we really be trying? *The New York Academy of Sciences*, 22 ago. 2024. Disponible en: https://www.nyas-org.translate.google.com/ideas-insights/blog/have-we-passed-the-turing-test-and-should-we-really-be-trying/?x_tr_sl=en&x_tr_tl=es&x_tr_hl=es&x_tr_pto=tc. Consultado el: 06 oct. 2025.

VISUS, Andrés. Que es un deep fakes, cómo se crean, cuáles fueron los primeros y su futuro. *ESIC University*, jul. 2021. Disponible en: <https://www.esic.edu/rethink/tecnologia/deep-fakes-que-es-como-se-crean-primeros-y-futuros>. Consultado el: 06 oct. 2025.

Data de submissão: 17 out. 2025
Data de aprovação: 05 dez. 2025